Using Repeated Measurements to Predict Cardiovascular Risk in Patients With Type 2 Diabetes Mellitus

Krishna M Gokhale, MSc[a,*], Joht Singh Chandan, PhD[a], Chris Sainsbury, MD[a], Peter Tino, PhD[b], Abd Tahrani, PhD[c], Konstantinos Toulis, PhD[a], and Krishnarajah Nirantharakumar, MD[a]

The QRISK cardiovascular disease (CVD) risk assessment model is not currently optimized for patients with type 2 diabetes mellitus (T2DM). We aim to identify if the abundantly available repeatedly measured data for patients with T2D improves the predictive capability of QRISK to support the decision-making process regarding CVD prevention in patients with T2DM. We identified patients with T2DM aged 25 to 85, not on statin treatment and without pre-existing CVD from the IQVIA Medical Research Data United Kingdom primary care database and then followed them up until the first diagnosis of CVD, ischemic heart disease, or stroke/transient ischemic attack. We included traditional, nontraditional risk factors and relevant treatments for our analysis. We then undertook a Cox's hazards model accounting for time-dependent covariates to estimate the hazard rates for each risk factor and calculated a 10-year risk score. Models were developed for males and females separately. We tested the performance of our models using validation data and calculated discrimination and calibration statistics. The study included 198,835 (180,143 male with 11,976 outcomes and 90,466 female with 8,258 outcomes) patients. The 10-year predicted survival probabilities for females was 0.87 (0.87 to 0.87), whereas the observed survival estimates from the Kaplan-Meier curve for all female models was 0.87 (0.86 to 0.87). The predicted and observed survival estimates for males were 0.84 (0.84 to 0.84) and 0.84 (0.83 to 0.84) respectively. The Harrell's C-index of all female models and all male models were 0.71 and 0.69 respectively. We found that including time-varying repeated measures, only mildly improved CVD risk prediction for T2DM patients in comparison to the current practice standard. We advocate for further research using time-varying data to identify if the involvement of further covariates may improve the accuracy of currently accepted prediction models. **(Am J Cardiol 2024;210:133−142)**

Keywords: cardiovascular risk, epidemiology, extended Cox-proportional hazards regression, repeatedly measured data, time-varying co-variates, type 2 diabetes

The prevalence of type 2 diabetes (T2DM) is increasing steadily and is a major public health burden.[1] Diabetes is in the top 10 causes of death globally, whereas cardiovascular disease (CVD) is the most common cause of death.[2] One-third of the T2DM patients are known to be at risk of developing CVD during their lifetime.[3] In the United Kingdom, the health care costs and indirect costs to the economy due to CVD are estimated at £9 billion and £19 billion respectively each year.[4] In developed nations like the United Kingdom many healthcare policies, clinical practice guidelines and prevention programmes exist to control and prevent CVD across the nation.[5] These guidance documents often refer to the multitude of prediction models used to identify patients who are at the risk of developing CVD in general population, few of which are derived from electronic health records or are relevant to patients with T2DM.[6]

In the United Kingdom, QRISK[7,8] is used for CVD risk assessment in patients with and without diabetes. The National Institute for Health and Care Excellence, recommends to offer a full formal risk assessment using the QRISK CVD risk assessment tool for any adult under the age of 85 who is suspected to be at risk of CVD development.[9] Whereas, the European Association for the study of Diabetes recommends the FRAMINGHAM[10] CVD risk prediction model and DECODE[11] to calculate risk score. Most of these models are built for general population and are sub-optimal when applied to those with T2DM.[12−14] The United Kingdom Prospective Diabetes Study risk engine[15] is one of the few CVD risk assessment models specific for T2DM population. However, in external validation United Kingdom Prospective Diabetes Study was found to have poor to moderate discriminative performance and poor calibration and therefore newer CVD risk prediction

[a]Institute of Applied Health Research, University of Birmingham, Birmingham, United Kingdom; [b]School of Computer Science, University of Birmingham, Birmingham, United Kingdom; and [c]Institute of Metabolism and Systems Research, University of Birmingham, Birmingham, United Kingdom. Manuscript received January 10, 2023; revised manuscript received and accepted October 1, 2023.

Funding: none.

See page 140 for Declaration of Competing Interest.

*Corresponding author: Tel: +44 (0)121 414 5643.

E-mail address: k.m.gokhale@bham.ac.uk (K.M. Gokhale).

https://doi.org/10.1016/j.amjcard.2023.10.008

models for T2DM patients are encouraged.[16,17] This large-scale, substandard performance of risk scores may be explained by the fact that several key predictors, assessed at baseline do change over the course of the pre-specified time frame and, along with them, so does the probability of CVD occurrence. This change remains elusive in the current, static approach, unless key time-dependent trajectories are identified and incorporated in the model construction. In the absence of accurate CVD prediction models, initiation of preventive interventions may remain sub-optimal in this important cohort. This challenge is not solely confined to the United Kingdom, as noted in systematic reviews, much of the current research in this space globally considers using single baseline measurements without accounting for variability.[13,18,19] Hence, there is a global need for further research to explore the impacts of such variability. Some efforts have been made to account for repeated measurement in CVD risk prediction such as when exploring the role of repeated blood pressure in the US study, The Atherosclerosis Risk in Communities (ARIC), among other global cohorts.[20−22] However, these are not undertaken in cohorts specific to T2DM.

The National Institute for Health and Care Excellence recommends health care professionals to follow patients with diabetes at regular intervals and to conduct annual reviews.[23] Hence, the availability of repeatedly measured data including biomarkers is abundant in electronic health records for patients with T2DM in the United Kingdom and allows us a unique opportunity to ascertain the importance of dynamic risk prediction using repeated measurements. In this study, we aimed to develop a CVD risk prediction model using Cox's extended proportional hazards model for time-dependent covariates for male and female patients with T2DM.

Methods

Study design

A retrospective cohort study from January 01, 1998 to December 31 2019.

Data source

We used IQVIA Medical Research Data (IMRD), which incorporates data from The Health Improvement Network (THIN), A Cegedim Database, a population-based primary care database. Reference made to THIN is intended to be descriptive of the data asset licensed by IQVIA. This work used de-identified data provided by patients as a part of their routine primary care. IMRD-UK contains 787 practices with over 15 million patient records. Information on demography, clinical diagnosis, symptoms, physical measurements, laboratory results, and prescriptions are available in the database. Clinical diagnosis and symptoms are stored using a hierarchical coding system (Read codes), whereas prescriptions are recorded using the Dictionary of Medicines and Devices (DM+D) classification system.[24,25] The database has been shown to be representative of the demographic structure of the United Kingdom population[26] and has been extensively used for diabetes mellitus and CVD research.[27−31] We defined an eligible period for each contributing general practice, as 12 months after either their acceptable mortality recording (AMR; a measure of data quality)[32] or installation of electronic health record after the study start date (January 01, 1998). A 1-year lag period was applied to ensure high-quality data. Data was extracted, transformed and loaded using DExtER.[33]

Study Population and study period

Our population of interest were patients with T2DM. We used Read codes to determine diagnoses of T2DM in eligible patients from January 01, 1998 to December 31, 2019. To be eligible patients had to be aged between 25 and 85 years at the time of cohort entry and registered with the general practice for at least 1 year. The index date was set to the day of diagnosis for newly diagnosed (incident patients) patients and study start date or the patient registration date plus 1 year (whichever was the latest), for patients who had a pre-existing diagnosis of T2DM (prevalent patients). Among the eligible patients we excluded anyone who had a recording of type 1 diabetes mellitus Read codes. Patients with a pre-existing code of the outcome of interest (ischemic heart disease [IHD] and stroke / transient ischemic attack [TIA]) at index date were excluded. If markers of pre-existing CVD, such as IHD, stroke, TIA, heart failure or relevant treatments, were present these patients were excluded from the cohort. We also excluded patients on statin at baseline as the main purpose of predicting CVD risk is to initiate statin treatment.[14] The cohort was followed up till earliest event of death, patient transferred out from practice, study end date, or the patient developed the outcome of interest (IHD, stroke, and TIA).

Outcome and Predictor variables

Our primary outcome of interest was the first diagnosis of IHD or Stroke and TIA after the index date. We referred to previous literature to determine appropriate risk factors for CVD in patients with T2DM.[6,8,18,34−36] We found many risk factors in our review of the literature, but we were able to include only those variables which were available in IMRD. Potential risk factors ranged from socio-demographic characteristics, lifestyle factors, medical history and diagnoses, physical measurements, and laboratory test results and prescriptions. Socio-demographic and lifestyle factors included age, sex, Townsend score, a measure of deprivation[37] (Ranging from 1 − least deprived, to 5 − most deprived), ethnicity, and smoking status. Medical history included family history of CVD, diagnosis of atrial fibrillation, hypothyroidism, rheumatoid arthritis, systemic lupus erythematosus, asthma, severe mental illness, anxiety, depression, hypertension, and migraine. In addition, we included complications from T2DM such as diabetic foot ulcer/amputation, peripheral neuropathy, retinopathy, and erectile dysfunction or treatment for erectile dysfunction in men determined by a diagnosis code, or at least one prescription of associated treatment (British National Formulary chapter 7.4.5).[38] Physical measurement and blood test results ranged from body mass index (BMI), systolic blood

pressure, blood lipid ratio, triglycerides, Albumin Creatinine ratio, Hemoglobin A1C (HbA1c), and estimated Glomerular Filtration Rate values, which were calculated from serum creatinine values using the chronic kidney disease epidemiology collaboration equation.[39] As studies show treating risk factors decreases the effect on the outcome[40,41]; treatments were included as potential predictors. Hence, we included insulin, angiotensin-converting enzyme inhibitors, and newly initiated statin treatment. We also included corticosteroid treatment based on previous risk prediction models.[6]

Derivation and validation of the models

Data preparation

Since our aim was to build a prediction model that uses time-dependent (repeatedly measured) data, we designed our dataset in a specific format. We calculated person years for each patient and then divided the person years into 6-month intervals for all patients. For each patient starting from first time interval, we looked for covariates at each 6-month interval until the end of follow up. For medical conditions, if a diagnosis code was found at baseline, then it was marked as present for all the time intervals. For cases where a patient develops a disease condition in the middle of the follow up, then that condition would be marked as present only for all the time intervals after the diagnosis. We treated drugs as categorical variables, and they were marked as present if there was at least one prescription recorded within the time interval; if none was found, it was marked as absent. There were cases where, for example, the smoking status of the patient changed within a time interval. In such cases, we looked at the status which was present for most part of the interval and retained that value. We treated ethnicity and family history of CVD as constants. For continuous variables such as BMI or blood pressure if more than one value was found within the interval, then the average was retained. Blood lipid ratio was not normally distributed, and it was log-transformed. To allow sufficient time for the medications to have an effect and to eliminate any reverse causality we lagged the time-varying data for each patient by one time interval (6 months). This meant we had to discard any patients where the follow-up period was less than or equal to 6 months.

In our data, the following predictors had missing information: smoking (21.6% missing), blood lipid ratio (35.2% missing), BMI (10.8% missing), systolic blood pressure (7.8% missing), HbA1c (18% missing), albumin creatinine ratio (42.1% missing), and triglycerides (27.3% missing). We used multiple imputation using chained equations (mice package in R)[42] to impute missing values and created 5 imputed datasets, we included all the predictor variables in the imputation model along with the Nelson-Allen estimator which improves imputation for left-censored data[43] and used Rubin's rules to combine the results from the imputed datasets. Missing data in Ethnicity and Townsend score were treated as separate class levels. We separated male and female datasets and for each we randomly allocated 70% of the data for derivation of the risk prediction models and the remaining 30% was used for internal validation.

Derivation of the models

We employed Cox's extended proportional hazards model for time-dependent covariates to estimate the hazard rate for each risk factor for all the models. We developed a model for females and males separately. We calculated a 10-year estimated risk of developing CVD.

Validation of the models

We tested the performance of all 4 models using the validation data and calculated discrimination and calibration. Missing data was filled in the same way as that of the development dataset. To evaluate the performance of the model, we calculated Harrell's C statistic[44] at 10 years which is an equivalent of the area under the receiver operating characteristics curve. The calibration of the model was assessed by plotting the observed survival probabilities (calculated using Kaplan-Meier estimates) against the predicted survival probabilities at 10 years.

We have followed the TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis) statement for reporting. All our experiments were conducted in R (version 3.5.2) and we mainly made use of rms,[45] survival,[46] and mice packages.

Sensitivity analyses

We have run 2 sensitivity analyses to assess the robustness and comparability of our findings:

Firstly, the current models "All Male" and "All Female" included both incident and prevalent diabetes cohorts, whereas we have also developed 2 additional models "Incident Male" and "Incident Female." These include only newly diagnosed T2DM patients.

Secondly, as one of the goals of the project is to compare whether our dynamic model improves the predictive capability of QRISK, we have also reported the performance of QRISK using our dataset and calculated the net reclassification index (NRI) to support comparison of the models. Full statistical details regarding the NRI are noted in online supplement 5.

Results

Study population

During the study, period we identified a total of 15,011,524 registered patients in the THIN database. After the application of the study exclusion criteria, 198,835 (1.3%) patients with a GP recorded code of T2DM were deemed eligible to form the exposed cohort. Among the 198,835 patients, 74,627 (37.6%) patients were identified as prevalent exposure (exposure code recorded before cohort entry), and the remaining 123,982 (62.4%) were newly diagnosed patients during the study period. Figure 1 describes the cohort selection flowchart.

Baseline characteristics

Of the total eligible cohort (198,835 patients), 90,466 (45.6%) were female, and 108,143 (54.4%) were male (further details of the cohort characteristics separated by sex can be found in Table 1).

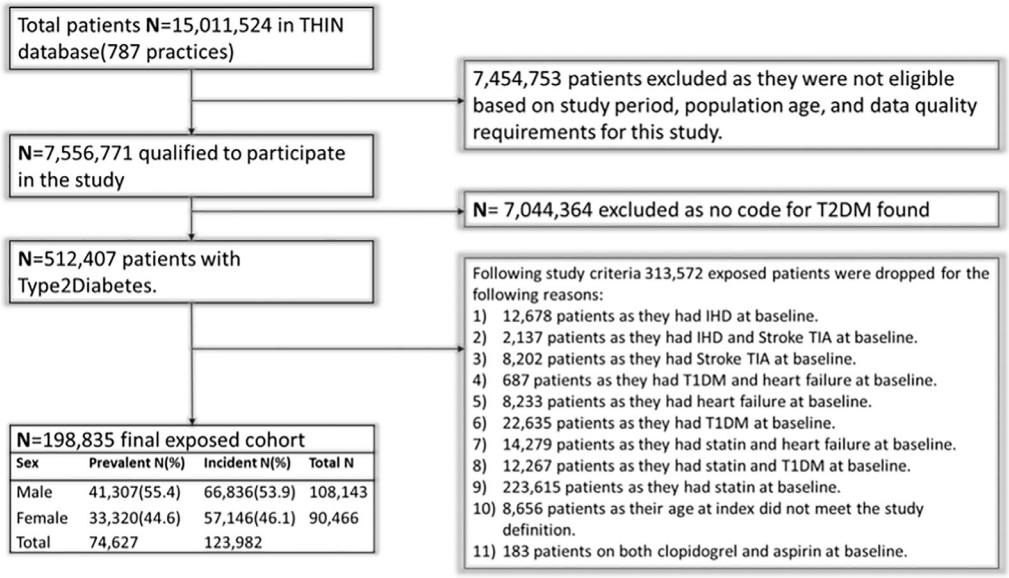

Figure 1. Cohort selection flowchart.

The median follow-up period for both men and women was similar (5 years). At study entry females were on average older than males (60 years compared with 58 years respectively) counterparts, and also experienced a higher proportion of deprivation. Ethnicity data was largely missing (54%) in both groups.

At study entry, the female cohort had a higher mean BMI (32.8 kg/m^2) compared with their male counterparts (30.7 kg/m^2) although they had lower rates of current smoking status (16.3% compared with 20.1%). Systolic blood pressure was similar between the groups. There were few differences between the groups when examining laboratory readings although there were slightly increased average levels of HbA1c (female: 58.1 mmol/L (SD 16.5); male: 59.6 mmol/L (SD 17.3) in males), blood lipids (female: 4.48 mmol/L (SD 2.34); male: 4.88 mmol/L (SD 2.46) and triglycerides (female: 1.88 mmol/L (SD 1.17); male: 2.04 mmol/L (1.57)) in males compared with females. In comparison, females had a higher proportion of kidney disease. There were similar rates of co-existing diabetes complications, comorbidities (although higher rates of hypertension, hypothyroidism, asthma, and mental ill health in the female cohort), and prescription medications between the groups.

Main findings

During the study period, a total of 20,414 (10.3%) of eligible patients developed the outcomes of interest (IHD/Stroke/TIA). Of these 12,302 (60%) and 8,112 (40%) developed IHD and stroke/ respectively. The female cohort accounted for 8,258 (IHD: 4,457; stroke/TIA: 3,801) of the total outcomes whereas males developed 12,156 (IHD: 4,311; stroke/TIA: 7,845) outcomes. Regardless of sex, the most substantial risk factors of developing CVD in the T2DM population are being of a South Asian risk factor, being a current smoker, having pre-existing diabetes complications, and also having treatment with insulin. Other factors including increasing age and deprivation, having a family history of CVD, having an increased systolic blood pressure, blood lipid ratio, triglycerides, reduced kidney function as well as pre-existing comorbidities and prescription medications were also all indicative of increasing risk of CVD in a T2DM population. Table 2 demonstrates the risk of CVD in the T2DM population when examined by each variable of interest broken by specific sex.

After the development of the predictor models demonstrated in Table 2, we were able to conduct and present the internally validated discrimination statistics for each of the models (Table 3). In the validation cohort for all females (in the study) the Harrell's C statistic (equivalent to the receiver operating curve for survival data) was 0.71, whereas it was 0.69 for males.

In the sensitivity analysis including incident only male and female patients, the Harrell's C statistic remained very similar (0.7 for incident only females, 0.67 for incident only males).

In order to calibrate the models, we plotted the probabilities against observed survival estimates which determined using the Kaplan-Meier estimator.[47]

Figure 2. visually demonstrates the Kaplan-Meier estimates for the patients included in the main analysis. The 10 year predicted survival probabilities for females was 0.87 (0.87 to 0.87), whereas the observed survival estimates from the Kaplan-Meier curve for all female models was 0.87 (0.86 to 0.87). The predicted and observed survival estimates for males were 0.84 (0.84 to 0.84) and 0.84 (0.83 to 0.84) respectively. All the models were well calibrated except for the incident only females where our model marginally overestimated.

Sensitivity analysis

During the sensitivity analysis including incident only patients (i.e., those patients who experienced the exposure of interest and were eligible to enter the study during the

Table 1
Baseline study characteristics

BASELINE STUDY CHARACTERISTICS			
		Female	Male
Total number of patients	n (%)	90466 (45.6)	108143 (54.4)
Prevalent patients	n (%)	33320 (36.8)	41307 (38.2)
Incident patients	n (%)	57146 (63.2)	66,836 (61.8)
Years of follow-up	Mean (SD), [Median]	5.58 (3.3), [5]	5.58 (3.3), [5]
DEMOGRAPHICS			
Age	Mean (SD), [Median]	59.4 (14.1), [60.0]	57.8 (12.9), [57.8]
Ethnicity			
White	n (%)	34486 (38.1%)	41777 (38.6%)
South Asian	n (%)	4011 (4.4%)	4267 (3.9%)
Black	n (%)	2160 (2.4%)	2150 (2.0%)
Mixed Race	n (%)	786 (0.9%)	843 (0.8%)
Other Ethnicity	n (%)	335 (0.4%)	330 (0.3%)
Missing	n (%)	48810 (53.9%)	58878 (54.4%)
Townsend			
1 (Least Deprived)	n (%)	15426 (17.0%)	21398 (19.8%)
2	n (%)	15404 (17.0%)	20001 (18.5%)
3	n (%)	17315 (19.1%)	20020 (18.5%)
4	n (%)	17297 (19.1%)	18591 (17.2%)
5	n (%)	13158 (14.5%)	13661 (12.6%)
Missing	n (%)	11988 (13.2%)	14574 (13.5%)
LIFESTYLE AND PHYSICAL MEASUREMENTS			
Body mass index (kg/m2)	Mean (SD), [Median]	32.8 (7.84), [31.6]	30.7 (6.17), [29.8]
	Missing n (%)	4077 (4.5%)	4671 (4.3%)
Smoking status			
Non-smoker	n (%)	54153 (59.8%)	48067 (44.4%)
Ex-smoker	n (%)	19999 (22.1%)	36441 (33.7%)
Smoker	n (%)	14742 (16.3%)	21750 (20.1%)
Missing	n (%)	1694 (1.9%)	1987 (1.8%)
Systolic blood pressure (mmHg)	Mean (SD), [Median]	140 (19.7), [140]	140 (18.4), [140]
	Missing n (%)	953 (1.1%)	1287 (1.2%)
LABAROTORY TEST RESULTS			
HbA1c (mmol/L)	Mean (SD), [Median]	58.1 (16.5), [54.1]	59.6 (17.3), [55.9]
	Missing n (%)	6311 (7.0%)	7529 (7.0%)
Blood lipid ratio (mmol/L)	Mean (SD), [Median]	4.48 (2.34), [4.10]	4.88 (2.46), [4.50]
	Missing n (%)	16320 (18.0%)	19526 (18.0%)
Triglycerides (mmol/L)	Mean (SD), [Median]	1.88 (1.17), [1.60]	2.04 (1.57), [1.65]
	Missing n (%)	39350 (43.4%)	46879 (43.3%)
Estimated glomerular filtration rate (EGFR) (ml/min/1.73m2)			
60+	n (%)	55555 (61.3%)	71185 (65.8%)
45-60	n (%)	8362 (9.2%)	6994 (6.5%)
30-45	n (%)	3259 (3.6%)	1371 (1.3%)
<30	n (%)	496 (0.5%)	365 (0.3%)
Missing	n (%)	22916 (25.3%)	28330 (26.2%)
Albumin creatinine ratio (mg/g)			
<3	n (%)	27043 (29.9%)	34370 (31.8%)
3 to 30	n (%)	7178 (7.9%)	8074 (7.5%)
30+	n (%)	648 (0.7%)	925 (0.9%)
Missing	n (%)	55719 (61.5%)	64876 (59.9%)
CO-EXISTING DIABETES COMPLICATIONS			
Retinopathy	n (%)	5352 (5.9%)	7192 (6.6%)
Peripheral Neuropathy	n (%)	647 (0.7%)	984 (0.9%)
Foot ulcer or Amputation	n (%)	1,810 (2.0%)	2,137 (2.0%)
Erectile Dysfunction or Erectile Dysfunction Treatment	n (%)	N/A	16865 (15.6%)

(*continued*)

Table 1 (*Continued*)

BASELINE STUDY CHARACTERISTICS			
		Female	Male
COMORBIDITIES			
Hypertension	n (%)	39635 (43.8%)	40786 (37.7%)
Hypothyroidism	n (%)	10084 (11.1%)	2421 (2.2%)
Atrial fibrillation	n (%)	2136 (2.4%)	2737 (2.5%)
Rheumatoid arthritis	n (%)	1486 (1.6%)	765 (0.7%)
Systemic lupus erythematous	n (%)	125 (0.1%)	24 (0.0%)
Migraine	n (%)	7028 (7.8%)	3038 (2.8%)
Asthma	n (%)	13155 (14.5%)	10237 (9.5%)
Severe mental illness	n (%)	3490 (3.9%)	2257 (2.1%)
Anxiety	n (%)	16657 (18.4%)	11327 (10.5%)
Depression	n (%)	22156 (24.5%)	14227 (13.1%)
Family history of CVD	n (%)	1321 (1.5%)	1469 (1.4%)
PRESCRIPTION MEDICATIONS			
Insulin	n (%)	4547 (5.0%)	4834 (4.5%)
Corticosteroids	n (%)	2946 (3.3%)	1854 (1.7%)
ACE-Inhibitors	n (%)	28753 (31.7%)	34114 (31.5%)

study period), the findings remained robust (Supplementary Tables 1, 2 and 3). When we replicated QRISK in our dataset we noted a c-index of 0.68 for men and 0.71 for women (Supplementary Table 4). After calculating the NRI the difference is negligible (−1), with some indication that the QRISK is mildly better at assigning patients to the correct risk categories.

Discussion

Here, we present one of the first cardiovascular risk prediction models that utilize repeatedly measured data for patients with T2DM. We found the magnitude and direction of our effect sizes from our models are consistent with the known literature, note the Harrell's C-index of all female models and all male models was 0.71 and 0.69 respectively. In comparison, the QRisk3 CVD risk assessment tool has a c-index of 0.70 for both men and women with T2DM.[48] Although in our dataset, QRISK has a c-index of 0.68 for men and 0.71 for women. Hence, compared with QRisk3 in their published validation dataset our models performed on par and mildly better in women, although in our dataset our model only demonstrates a mild improvement for men. Another example is the diabetes severity score which was developed using United Kingdom primary and secondary care linked data.[49] The primary outcome in this scoring system was mortality. However, CVD hospitalization was a secondary outcome and the model performed reasonably with an area under the receiver operating characteristics=0.73 to the describe the association between a one-unit score and the risk of hospitalization. However, this model did not include the breadth of CVD which is not-hospitalized, and hence it is difficult to compare the findings accordingly.

With regards to the clinical implications of the present study, we argue that our model, as well as the existing models designed to identify those patients with diabetes in whom intervention for primary CVD prevention is warranted, do not meet rigorous clinical standards and their routine used to guide therapy initiation or intensity of intervention should be applied with caution and considered to have ancillary gravity in clinical decision process. Moderate intensity statin administration for primary CVD prevention in patients with diabetes, 40 to 75 years of age, without calculating any CVD risk score, presents an alternative approach and forms a current recommendation from the American Heart Association.[50]

Whether incorporating repeatedly measured data should be the new norm or a right step in the direction of optimizing CVD risk assessment model for patients with diabetes is yet unclear. Our research has allowed us to build on previous studies which explore the use of repeated measurements in CVD risk prediction.[20–22,51] However, we note that as described previously in the literature, the use of repeated measurements appears to only lead to a slight to modest improvement in the overall risk prediction.[52] A recent study by Xu et al[53] also tried to answer a similar question using an alternative United Kingdom primary care dataset. Our results are most comparable to the methodology that they employed which also demonstrates a moderate improvement when accounting for variability of measurements in this particular cohort. However, the added value of this approach may still not be clinically important.

Although our study findings should be interpreted within the context of its limitations, the use of repeatedly measured data resulted in an incremental, minor improvement in the performance of the model over the currently accepted practice. The magnitude of this incremental improvement may be regarded as of limited clinical impact, adds to the uncertainty regarding CVD risk score in the context of T2DM and points to potential alternative explanation of the findings. For example, genetic predisposition could be a key predictor which is missing from current candidate covariate list. The same applies to the epigenetic and environmental

Table 2
Risk of cardiovascular disease

Variable	Male (Hazard ratio; 95% Confidence intervals)	Female (Hazard ratio; 95% Confidence intervals)
Demography		
Age	1.04 (1.04 - 1.04)	1.04 (1.04 - 1.05)
Townsend (ref= 1 least deprived)		
2	1.07 (1 - 1.14)	1.13 (1.04 - 1.24)
3	1.14 (1.07 - 1.22)	1.14 (1.04 - 1.24)
4	1.21 (1.12 - 1.29)	1.23 (1.13 - 1.34)
5 (most deprived)	1.24 (1.15 - 1.34)	1.4 (1.28 - 1.53)
Missing	1.11 (1.03 - 1.2)	1.04 (0.94 - 1.15)
Ethnicity (ref= White)		
South Asian	1.67 (1.49 - 1.87)	1.45 (1.25 - 1.68)
Black	0.67 (0.54 - 0.84)	0.87 (0.69 - 1.1)
Mixed Race	1.27 (0.95 - 1.68)	1.16 (0.79 - 1.71)
Others (Includes Chinese, Middle Eastern)	0.88 (0.52 - 1.48)	0.72 (0.38 - 1.4)
Missing	1.17 (1.12 - 1.23)	1.23 (1.17 - 1.3)
Family History of CVD	1.52 (1.28 - 1.8)	1.53 (1.26 - 1.86)
Lifestyle and physical measurements		
Smoking (ref= non-smoker)		
Ex-smoker	1.03 (0.98 - 1.08)	1.06 (0.99 - 1.12)
Current Smoker	1.25 (1.15 - 1.34)	1.39 (1.27 - 1.52)
Body mass index (kg/m2)	0.99 (0.99 - 1)	1 (0.99 - 1)
Systolic Blood Pressure	1.01 (1.01 - 1.01)	1.01 (1.01 - 1.01)
Laboratory test results		
HbA1C	1 (1 - 1)	1 (1 - 1)
Blood lipid ratio	1.03 (1.01 - 1.05)	1.03 (1.01 - 1.05)
Triglycerides	1.02 (1.01 - 1.04)	1.05 (1.02 - 1.07)
Estimated Glomerular Filtration Rate (eGFR) (ref= 60+)		
45 to 60	1.09 (1.01 - 1.18)	1.06 (0.97 - 1.15)
30 to 45	1.17 (1.06 - 1.29)	1.16 (1.07 - 1.26)
<30	1.15 (0.98 - 1.35)	1.29 (1.1 - 1.51)
Albumin Creatinine ratio (ref= <3)		
3 to 30	1.06 (1 - 1.11)	1.09 (1.03 - 1.16)
30+	1.22 (1.09 - 1.35)	1.42 (1.22 - 1.64)
Diabetes complications		
Retinopathy	1.23 (1.17 - 1.3)	1.24 (1.16 - 1.32)
Peripheral Neuropathy	1.28 (1.13 - 1.45)	1.14 (0.97 - 1.33)
Diabetic Foot Ulcer or Amputation	1.33 (1.22 - 1.46)	1.4 (1.26 - 1.55)
Erectile Dysfunction or Erectile Dysfunction Treatment	1.15 (1.1 - 1.21)	(-)
Comorbidities		
Hypertension	1.09 (1.03 - 1.15)	1.1 (1.03 - 1.17)
Hypothyroidism	0.97 (0.87 - 1.08)	0.95 (0.89 - 1.03)
Atrial fibrillation	1.38 (1.26 - 1.5)	1.65 (1.5 - 1.82)
Rheumatoid arthritis	1.43 (1.2 - 1.7)	1.31 (1.11 - 1.53)
Lupus SLE	1.52 (0.38 - 6.11)	1.02 (0.51 - 2.06)
Migraine	1.18 (1.04 - 1.34)	1.17 (1.05 - 1.29)
Asthma	1.06 (0.98 - 1.13)	1.14 (1.07 - 1.23)
Severe mental illness	0.93 (0.8 - 1.08)	0.98 (0.85 - 1.13)
Anxiety	1.05 (0.97 - 1.12)	1.01 (0.94 - 1.09)
Depression	1.18 (1.1 - 1.26)	1.16 (1.09 - 1.24)
Prescription medications		
Corticosteroids	1.16 (0.99 - 1.36)	1.03 (0.9 - 1.18)
Insulin	1.32 (1.24 - 1.41)	1.41 (1.31 - 1.52)
Statins	0.9 (0.86 - 0.94)	0.85 (0.8 - 0.9)
ACE-Inhibitors	0.96 (0.91 - 1.01)	1.07 (1 - 1.13)

Table 3
Discrimination statistics

	All Male	All Female
Harrell's C-index	0.69	0.71

influences, neither of which are recorded well in primary care data. Similarly, the absence of newer biomarkers, such as pro-BNP, from the candidate covariate list maybe a limitation.

We advocate for future research incorporating such temporal information in their model building, while also ascertaining the use of more sophisticated statistical or machine learning techniques, which might be preferable when pattern recognition or both linear and nonlinear relationships are involved. Recently, the superiority of deep learning techniques for survival analysis over traditional Cox models has been reported in the literature. This would be a future research priority that might offer the next step toward optimizing CVD model prediction.

In conclusion, we developed a dynamic CVD risk prediction model for T2DM patients that utilized repeatedly measured data. Although our models perform mildly better than current state of the art, it is not practice changing and as such, we support further research in this space particularly using explainable deep learning models to improve risk prediction.

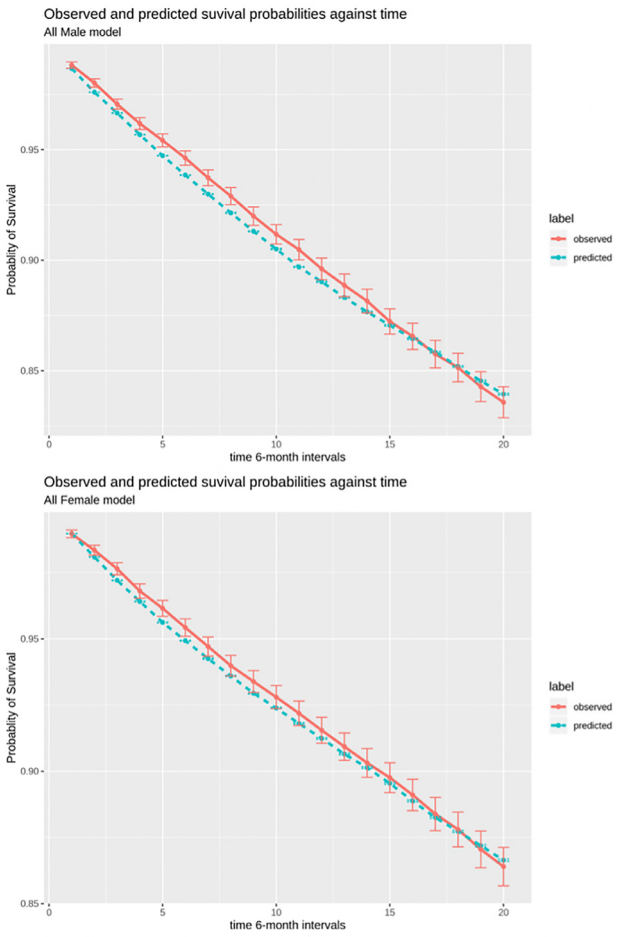

Figure 2. Kaplan-Meier estimates.

Declaration of Competing Interest

Drs. Chandan, Nirantharakumar, and Gokhale are codirectors of data extraction for epidemiological research (Dexter) operating division which is part of the University of Birmingham. Dexter operating division supports the extraction and preparing of healthcare data to support epidemiological analyses such as those seen in this article. The remaining authors have no competing interests to declare.

Authors' Contributions

Conceptualization (Krishna M. Gokhale, Krishnarajah Nirantharakumar), data curation and verification of underlying data (Krishna M. Gokhale, Krishnarajah Nirantharakumar), data analysis and access to data (Krishna M. Gokhale), methodology (Krishna M. Gokhale, Krishnarajah Nirantharakumar, Peter Tino), writing - original draft (Krishna M. Gokhale, Joht Singh Chandan, Chris Sainsbury, Peter Tino, Abd Tahrani, Konstantinos Toulis, Krishnarajah Nirantharakumar), writing - review & editing (Krishna M. Gokhale, Joht Singh Chandan, Chris Sainsbury, Peter Tino, Abd Tahrani, Konstantinos Toulis, Krishnarajah Nirantharakumar).

Ethical Approval

Anonymized data was used from the data provider to the University of Birmingham. The use of IQVIA Medical Research Data is approved by the United Kingdom Research Ethics Committee (reference number: 18/LO/0441); in accordance with this approval, the study protocol must be reviewed and approved by an independent Scientific Review Committee (SRC). The protocol has been approved for this project by the independent SRC (19THIN018). Specifically for the analysis undertaken in this manuscript, an amendment was made to allow for the exploration of atopic disorders as the previous protocol specified cardiometabolic, mental health and central sensitization outcomes. IQVIA Medical Research Data incorporates data from The Health Improvement Network (THIN), a Cegedim Database. Reference made to THIN is intended to be descriptive of the data asset licensed by IQVIA. This work has used deidentified data provided by patients as a part of their routine primary care. As the data is de-identified there is no opportunity/ability for the research team to seek independent written consent from those who contribute to the dataset.

Data Sharing

Upon publication, the analysis code will be available upon request from the corresponding author (Dr. Gokhale). In order to obtain access to the raw data that was used for all the analyses, approval must be sought from the data provider (IQVIA) and independent SRC who approved the ethics for this project who will be able to share the IQVIA Medical Research Data dataset. It is likely that this process

will incur a cost bespoke to the Institution requesting the data. This can be done with support of the corresponding author (Dr. Gokhale).

Supplementary materials

Supplementary material associated with this article can be found in the online version at https://doi.org/10.1016/j.amjcard.2023.10.008.

1. Zheng Y, Ley SH, Hu FB. Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. *Nat Rev Endocrinol* 2018;14:88–98.
2. Murray JL. Global , regional , and national age − sex specifi c all-cause and cause-specifi c mortality for 240 causes of death , 1990 −2013 : a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* 2013;385:1990–2013.
3. Einarson TR, Acs A, Ludwig C, Panton UH. Economic burden of cardiovascular disease in type 2 diabetes: a systematic review. *Value Health* 2018;21:881–890.
4. Nichols M, Townsend N, Nichols M, Townsend N, Luengo-Fernandez R, Leal J, Gray A, Scarborough P, Rayner M. *European Cardiovascular Disease Statistics 2012 Edition*. Brussels: European Heart Network; 2012.
5. National Institute for Health and Care Excellence. Cardiovascular disease: risk assessment and reduction, including lipid modification. Available at: https://www.nice.org.uk/guidance/cg181. Accessed on January 8, 2023.
6. Damen JAAG, Hooft L, Schuit E, Debray TPA, Collins GS, Tzoulaki I, Lassale CM, Siontis GC, Chiocchia V, Roberts C, Schlüssel MM, Gerry S, Black JA, Heus P, van der Schouw YT, Peelen LM, Moons KG. Prediction models for cardiovascular disease risk in the general population: systematic review. *BMJ* 2016;353:i2416.
7. Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, Minhas R, Sheikh A, Brindle P. Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2. *BMJ* 2008;336:1475–1482.
8. Hippisley-Cox J, Coupland C. Development and validation of QDiabetes-2018 risk prediction algorithm to estimate future risk of type 2 diabetes: cohort study. *BMJ* 2017;359:j5019.
9. Rabar S, Harker M, O'flynn N, Wierzbicki AS, Guideline Development Group. Lipid modification and cardiovascular risk assessment for the primary and secondary prevention of cardiovascular disease: summary of updated NICE guidance. *BMJ* 2014;349:g4356.
10. Eichler K, Puhan MA, Steurer J, Bachmann LM. Prediction of first coronary events with the Framingham score: a systematic review. *Am Heart J* 2007;153:722–731.e8.
11. Balkau B, Hu G, Qiao Q, Tuomilehto J, Borch-Johnsen K, Pyörälä K, DECODE Study Group, European Diabetes Epidemiology Group. Prediction of the risk of cardiovascular mortality using a score that includes glucose as a risk factor. The DECODE study. *Diabetologia* 2004;47:2118–2128.
12. Brindle P, Beswick A, Fahey T, Ebrahim S. Accuracy and impact of risk assessment in the primary prevention of cardiovascular disease: a systematic review. *Heart* 2006;92:1752–1759.
13. Chamnan P, Simmons RK, Sharp SJ, Griffin SJ, Wareham NJ. Cardiovascular risk assessment scores for people with diabetes: a systematic review. *Diabetologia* 2009;52:2001–2014.
14. Read SH, van Diepen M, Colhoun HM, Halbesma N, Lindsay RS, McKnight JA, McAllister DA, Pearson ER, Petrie JR, Philip S, Sattar N, Woodward M, Wild SH, Scottish Diabetes Research Network Epidemiology Group. Performance of cardiovascular disease risk scores in people diagnosed with type 2 diabetes: external validation using data from the national Scottish diabetes register. *Diabetes Care* 2018;41:2010–2018.
15. Stevens RJ, Kothari V, Adler AI, Stratton IM, United Kingdom Prospective Diabetes Study (UKPDS) Group. The UKPDS risk engine: A model for the risk of coronary heart disease in type II diabetes (UKPDS 56) [published correction appears in *Clin Sci (Lond)* 2002;102:679]. *Clin Sci (Lond)* 2001;101:671–679.
16. Van Dieren S, Peelen LM, Nöthlings U, Van der Schouw YT, Rutten GE, Spijkerman AMW, van der A DL, Sluik D, Boeing H, Moons KG, Beulens JW. External validation of the UK Prospective Diabetes Study (UKPDS) risk engine in patients with type 2 diabetes. *Diabetologia* 2011;54:264–270.
17. Bannister CA, Poole CD, Jenkins-Jones S, Morgan CL, Elwyn G, Spasić I, Currie CJ. External validation of the UKPDS risk engine in incident type 2 diabetes: a need for new type 2 diabetes−specific risk equations. *Diabetes Care* 2014;37:537–545.
18. Van Dieren S, Beulens JWJ, Kengne AP, Peelen LM, Rutten GE, Woodward M, van der Schouw YT, Moons KG. Prediction models for the risk of cardiovascular disease in patients with type 2 diabetes: a systematic review. *Heart* 2012;98:360–369.
19. Chowdhury MZI, Yeasmin F, Rabi DM, Ronksley PE, Turin TC. Prognostic tools for cardiovascular disease in patients with type 2 diabetes: a systematic review and meta-analysis of C-statistics. *J Diabetes Complications* 2019;33:98–111.
20. Sweeting MJ, Barrett JK, Thompson SG, Wood AM. The use of repeated blood pressure measures for cardiovascular risk prediction: a comparison of statistical models in the ARIC study. *Stat Med* 2017;36:4514–4528.
21. Paige E, Barrett J, Stevens D, Keogh RH, Sweeting MJ, Nazareth I, Petersen I, Wood AM. Landmark models for optimizing the use of repeated measurements of risk factors in electronic health records to predict future disease risk. *Am J Epidemiol* 2018;187:1530–1538.
22. Paige E, Barrett J, Pennells L, Sweeting M, Willeit P, Di Angelantonio E, Gudnason V, Nordestgaard BG, Psaty BM, Goldbourt U, Best LG, Assmann G, Salonen JT, Nietert PJ, Verschuren WMM, Brunner EJ, Kronmal RA, Salomaa V, Bakker SJL, Dagenais GR, Sato S, Jansson JH, Willeit J, Onat A, de la Cámara AG, Roussel R, Völzke H, Dankner R, Tipping RW, Meade TW, Donfrancesco C, Kuller LH, Peters A, Gallacher J, Kromhout D, Iso H, Knuiman M, Casiglia E, Kavousi M, Palmieri L, Sundström J, Davis BR, Njølstad I, Couper D, Danesh J, Thompson SG, Wood A. Use of repeated blood pressure and cholesterol measurements to improve cardiovascular disease risk prediction: an individual-participant-data meta-analysis. *Am J Epidemiol* 2017;186:899–907.
23. National Institute for Health and Care Excellence. Type 2 diabetes in adults: management. Available at: https://www.nice.org.uk/guidance/ng28. Accessed on January 8, 2023.
24. NHS Business Services Authority. Dictionary of medicines and devices (dm+d). Available at: https://www.nhsbsa.nhs.uk/pharmacies-gp-practices-and-appliance-contractors/dictionary-medicines-and-devices-dmd. Accessed on January 8, 2023.
25. Booth N. What are the Read codes? *Health Libr Rev* 1994;11:177–182.
26. Blak BT, Thompson M, Dattani H, Bourke A. Generalisability of the Health Improvement Network (THIN) database: demographics, chronic disease prevalence and mortality rates. *Inform Prim Care* 2011;19:251–255.
27. Toulis KA, Willis BH, Marshall T, Kumarendran B, Gokhale K, Ghosh S, Thomas GN, Cheng KK, Narendran P, Hanif W, Nirantharakumar K. All-cause mortality in patients with diabetes under treatment with dapagliflozin: A population-based, open-cohort study in the health improvement network database. *J Clin Endocrinol Metab* 2017;102:1719–1725.
28. Nirantharakumar K, Willis BH, Marshall T, Kumarendran B, Ghosh S, Narendran P, Thomas GN, Ponnusamy S, Hanif W, Toulis KA, Cheng K. *Glucagon-like peptide 1 agonists and death from any cause in primary care setting*. Diabetic Medicine. Hoboken: Wiley; 2017:159.
29. Toulis KA, Willis BH, Marshall T, Ghosh S, Gokhale K, Narendran P, Thomas GN, Cheng KK, Hanif W, Nirantharakumar K, Kumarendran B. *Sodium-glucose cotransporter 2 inhibitiors and death from any cause in a real world setting*. Diabetic Medicine. Hoboken: Wiley; 2017:160.
30. Chandan JS, Thomas T, Bradbury-Jones C, Taylor J, Bandyopadhyay S, Nirantharakumar K. Risk of cardiometabolic disease and all-cause mortality in female survivors of domestic abuse. *J Am Heart Assoc* 2020;9:e014580.
31. Chandan JS, Okoth K, Gokhale KM, Bandyopadhyay S, Taylor J, Nirantharakumar K. Increased cardiometabolic and mortality risk following childhood maltreatment in the United Kingdom. *J Am Heart Assoc* 2020;9:e015855.
32. Maguire A, Blak BT, Thompson M. The importance of defining periods of complete mortality reporting for research using automated data from primary care. *Pharmacoepidemiol Drug Saf* 2009;18:76–83.
33. Gokhale KM, Chandan JS, Toulis K, Gkoutos G, Tino P, Nirantharakumar K. Data extraction for epidemiological research (DExtER): a novel tool for automated clinical epidemiology studies. *Eur J Epidemiol* 2021;36:165–178.

34. Wang CCL, Hess CN, Hiatt WR, Goldfine AB. Atherosclerotic cardiovascular disease and heart failure in type 2 diabetes − mechanisms, management, and clinical considerations. *Circulation* 2016;133:2459.
35. Fox CS, Golden SH, Anderson C, Bray GA, Burke LE, De Boer IH, Deedwania P, Eckel RH, Ershow AG, Fradkin J, Inzucchi SE, Kosiborod M, Nelson RG, Patel MJ, Pignone M, Quinn L, Schauer PR, Selvin E, Vafiadis DK, American Heart Association Diabetes Committee of the Council on Lifestyle and Cardiometabolic Health, Council on Clinical Cardiology, Council on Cardiovascular and stroke Nursing, Council on Cardiovascular Surgery and Anesthesia, Council on Quality of Care and Outcomes Research, American Diabetes Association. Update on prevention of cardiovascular disease in adults with type 2 diabetes mellitus in light of recent evidence: a scientific statement from the American Heart Association and the American Diabetes Association. *Diabetes Care* 2015;38:1777–1803.
36. Narain R, Saxena S, Goyal AK. Cardiovascular risk prediction: a comparative study of Framingham and quantum neural network based approach. *Patient Prefer Adherence* 2016;10:1259–1270.
37. Townsend P, Phillimore P, Beattie A. *Health and Deprivation: Inequality and the North*. Oxfordshire: Taylor & Francis; 2023.
38. British Nutrition Foundation 7.4.5. Drugs for Erectile Dysfunction | OpenPrescribing. https://openprescribing.net/bnf/070405/. Accessed on January 8, 2023.
39. Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF, Feldman HI, Kusek JW, Eggers P, Van Lente F, Greene T, Coresh J, CKD-EPI (Chronic Kidney Disease Epidemiology Collaboration). A new equation to estimate glomerular filtration rate. *Ann Intern Med* 2009;150:604–612.
40. Wilson PW, Anderson KM, Castelli WP. Twelve-year incidence of coronary heart disease in middle-aged adults during the era of hypertensive therapy: the Framingham offspring study. *Am J Med* 1991;90:11–16.
41. Abdul-Ghani M, Del Prato S, Chilton R, DeFronzo RA. SGLT2 inhibitors and cardiovascular risk: lessons learned from the EMPA-REG Outcome study. *Diabetes Care* 2016;39:717–725.
42. van Buuren Sv, Groothuis-Oudshoorn K. mice: multivariate imputation by chained equations in R. *J Stat Softw* 2011;45:1–67.
43. White IR, Royston P. Imputing missing covariate values for the Cox model. *Stat Med* 2009;28:1982–1998.
44. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;15:361–387.
45. Harrell FE Jr. Package 'rms. Vanderbilt Univ; 2019. https://cran.r-project.org/web/packages/rms/rms.pdf. Accessed on January 8, 2023.
46. Therneau TM, Lumley T. Package 'Survival.' R Top Doc. 2015. https://cran.r-project.org/web/packages/survival/survival.pdf. Accessed on January 8, 2023.
47. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc* 1958;53:457–481.
48. Hippisley-Cox J, Coupland C, Brindle P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. *BMJ* 2017;357:j2099.
49. Zghebi SS, Mamas MA, Ashcroft DM, Salisbury C, Mallen CD, Chew-Graham CA, Reeves D, Van Marwijk H, Qureshi N, Weng S, Holt T, Buchan I, Peek N, Giles S, Rutter MK, Kontopantelis E. Development and validation of the DIabetes Severity SCOre (DISSCO) in 139 626 individuals with type 2 diabetes: a retrospective cohort study. *BMJ Open Diabetes Res Care* 2020;8:e000962.
50. Grundy SM, Stone NJ, Bailey AL, Beam C, Birtcher KK, Blumenthal RS, Braun LT, de Ferranti S, Faiella-Tommasino J, Forman DE, Goldberg R, Heidenreich PA, Hlatky MA, Jones DW, Lloyd-Jones D, Lopez-Pajares N, Ndumele CE, Orringer CE, Peralta CA, Saseen JJ, Smith SC, Sperling L, Virani SS, Yeboah J. 2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA guideline on the management of blood cholesterol: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice guidelines. *Circulation* 2019;139:e1082–e1143.
51. Ayala Solares JR, Canoy D, Raimondi FED, Zhu Y, Hassaine A, Salimi-Khorshidi G, Tran J, Copland E, Zottoli M, Pinho-Gomes AC, Nazarzadeh M, Rahimi K. Long-term exposure to elevated systolic blood pressure in predicting incident cardiovascular disease: evidence from large-scale routine electronic health records. *J Am Heart Assoc* 2019;8:e012129.
52. Goldstein BA, Pomann GM, Winkelmayer WC, Pencina MJ. A comparison of risk prediction methods using repeated observations: an application to electronic health records for hemodialysis. *Stat Med* 2017;36:2750–2763.
53. Xu Z, Arnold M, Sun L, Stevens D, Chung R, Ip S, Barrett J, Kaptoge S, Pennells L, Di Angelantonio E, Wood AM. Incremental value of risk factor variability for cardiovascular risk prediction in individuals with type 2 diabetes: results from UK primary care electronic health records. *Int J Epidemiol* 2022;51:1813–1823.